The authors would like to thank the editor, the two reviewers, and Dr. Ouimette for their thoughtful and thorough review, and constructive remarks. We have modified the manuscript based on these comments to improve and clarify the text. Please find below detailed responses in bold blue text (with direct quotes from the revised manuscript shown in "bold, quoted and italic" text) to the comments and suggestions offered by the reviewers (shown in normal text). All line numbers in our responses correspond to the "clean" version of the revised manuscript.

RESPONSE TO THE COMMENTS FROM REFEREE 1

This paper looks for a better PurpleAir correction for sensors in the US southeast. However, they only consider one equation from the literature when much additional work has been done on this topic in the past 4 years. This is not the first paper to look at nonlinear RH correction and the paper would be strengthened by comparing to other corrections in the literature that account for nonlinear RH. I have a number of other specific comments below that I hope the authors will address to strengthen their paper. The editor should also find someone to review that is more familiar with semi-supervised clustering.

Major

1- I think this paper would be strengthened by considering other common corrections from the literature especially those that consider nonlinear RH terms (e.g., Wallace https://www.mdpi.com/1424-8220/22/13/4741, Nilson https://amt.copernicus.org/articles/15/3315/2022/amt-15-3315-2022.html, Malings https://www.tandfonline.com/doi/full/10.1080/02786826.2019.1623863)

Response: The authors appreciate the reviewer's suggestion. We added a new paragraph (lines 369-377 in the Results and Discussion section) to compare the models developed in this study with other existing non-linear models as suggested. However, these models were designed for specific locations and not intended to work for a broad area. Moreover, none of these studies covered the Southeastern U.S. Malings et al. (2020) used data from 2 sites in Pittsburgh. Wallace et al. (2022) used data from California, Washington and Oregon. We added the results found by Wallace et al. (2001, 2022) and Malings et al. (2020). However, we did not include Nilson et al. (2022) since they only developed linear models using CF-1 PurpleAir data.

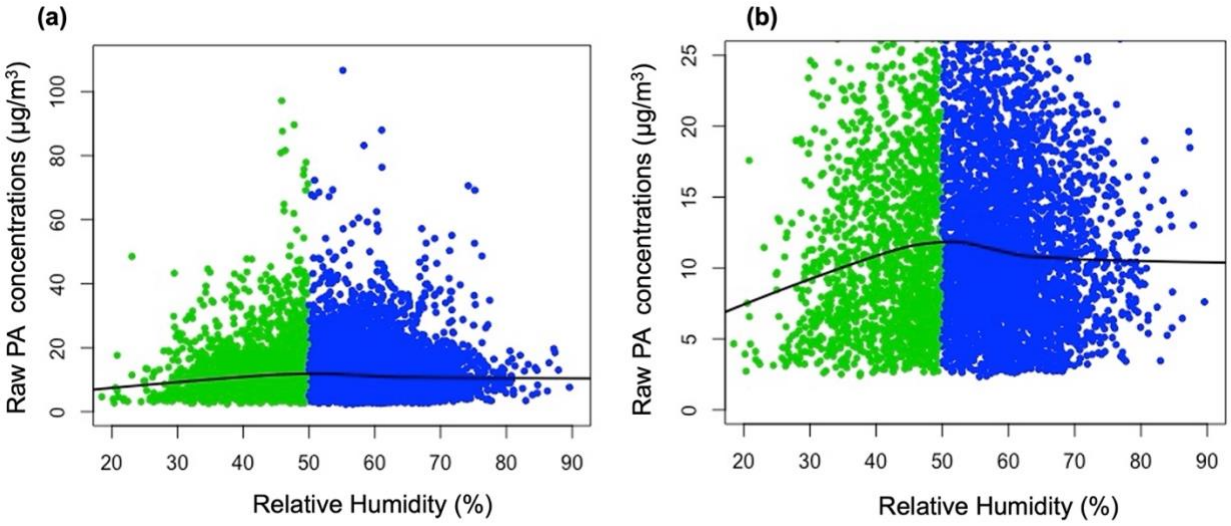
"We compared our results with nonlinear models that were previously developed and tested for PurpleAir sensor bias correction. Malings et al. (2020) developed a two-piece linear model based on a threshold of 20 µg m-3 PM2.5 concentrations using 11 PurpleAir sensors in 2 sites in Pittsburgh. The models included CPA, T, RH and DP as predictors. They found a correlation below 50 % and a MAE ranging from 3 to 5 µg m-3 (Malings et al., 2020). Some other studies (Wallace et al., 2021, 2022) estimated correction factors based on the ratio of the mean AQS to the mean PurpleAir for all pairs of PurpleAir/AQS sites first using 33 PurpleAir sensors from California (Wallace et al., 2021) and then including 182 PurpleAir sensors from

California, Washington and Oregon (Wallace et al., 2022). Their studies evaluated alternative PM2.5 PurpleAir estimates, however Wallace et al. (2021) also developed a correction factor for the cf_1 PM2.5 PurpleAir estimates. They calculated a range of a correction factors between 0.65 and 0.72 resulting in an overestimation of PM2.5 of 40 % compared with AQS monitors (Wallace et al., 2021).” **(lines 369-377)**

2- Also, can you add a plot showing the RH nonlinearity? You say that the model shows that it shows up around 50% but where does it increase visually? Something like RH on the X axis and Sensor/Monitor on the Y axis (Examples: Zheng https://amt.copernicus.org/articles/11/4823/2018/)

Response: We appreciate the suggestion. The plot has been added to the Supplemental Information (Figure S4) and referenced in the manuscript in line 342. Figure S4 shows the correlation between raw $PM_{2.5}$ PurpleAir concentrations and RH, with the regression line displaying the nonlinearity of $PM_{2.5}$ PurpleAir concentrations. The non-linearity curve started around RH = 50%.

“Figure S4 shows that non-linearity in the curve started around RH of 50%. PurpleAir datapoints that fell within a range of RH less or equal to 50% are in green and those that fell within a range greater than 50 % are shown in blue.”



“Figure S4: Correlation between raw $PM_{2.5}$ PurpleAir concentrations and RH showing the nonlinearity of $PM_{2.5}$ PurpleAir concentrations. Graph a) shows all the datapoints, and graph b) is a zoom in to better display the regression line and the nonlinearity of the data.”

3- This paper discusses how the southeast is unique because it is high humidity but it would also be helpful to comment on how particle properties (e.g., composition, size distribution) are different in the south east and how that might impact the performance (e.g., Patel https://amt.copernicus.org/articles/17/1051/2024/, Jaffe https://amt.copernicus.org/articles/16/1311/2023/).

Response: We thank the reviewer for the suggestion. We edited the manuscript to highlight specific sources of $PM_{2.5}$ and potential impact of particle properties in the Southeast region (lines 102-106).

***"The high humidity condition in this part of the U.S. might affect particle composition and size distribution due to water uptake (Hagan & Kroll, 2020; Jaffe et al., 2023; Patel et al., 2024; Rueda et al., 2023). A study conducted in 2018 (Carlton et al., 2018) found large contributions (50%) to $PM_{2.5}$ from biogenic secondary organic aerosols (BSOA) in the Southeast U.S. region compared with the rest of the country. The elevated BSOA are attributed to heavily forested areas and large urban areas in the region (U.S. EPA, 2018; Carlton et al., 2018)."* (lines 102-106)**

4- How does the recent release of the T640 correction impact this work? I agree with Dr. Ouimette that it would be helpful to list all the AQS monitors compared to, I assume some of them are Teledyne T640s.

Response: We thank the reviewer for the comment. We edited the manuscript to include limitations related to T640s in lines 287-289. The AQS monitors are listed in Fig. S13.

***"Similarly, the presence of Teledyne T640s among our AQS monitors may have affected the performance of our models since positive bias of approximately 20% has been reported with T640s compared with other FEM or FRM monitors (U.S. EPA, 2024)."* (lines 287-289)**

5- Were any of these sensors the alternate PMS5003s? Sear, Kaur, Kelly, https://www.sciencedirect.com/science/article/pii/S0021850223001210 How does this impact your results?

Response: We thank the reviewer for the comment. We edited the manuscript to include limitations related to the alternative PMS5003 in lines 290-295.

***"Additionally, a study conducted by Searle et al. (2023) found that 12.9 % of the sensors deployed by PurpleAir between June 2021 and May 2023 reported negative bias of approximatively 3 $\mu g\ m^{-3}$ over the long term. These PurpleAir sensors, specifically deployed between June 2021 and January 2022, and between March to May 2023 used an alternative, Plantower PMS5003 that affected the reported particle size distributions and concentrations (Searle et al., 2023). Although only 5 of our sensors, representing about 7 % of our data, fell into the reported time periods (Fig. 2), the potential presence of the alternative PMS5003 in our study may have affected the performance of our models."* (lines 290-295)**

6- How much data is excluded for each of the QA methods? (AB channel comparison high, low, etc.)

Response: The amount of data removed at each step of the QA process was estimated, and a table with this information was added in the Supplemental Information (Table S1) and referenced in the manuscript in line 236.

***"The QA process removed about 22 % (Table S1) of the raw data…"* (line 236)**

"Table S1: Percentage of hourly data removed by QA process from the initial 56 PurpleAir sensors

QA criteria	*% removed**
Process 1: Removing NAs (PM, T, RH)	*2.026*
Process 2: Channels A & B agreement *Low concentration (≤ 25 µg/m³): 537,246 obs.* *High concentration (>25 µg/m³): 80,196 obs.*	 *2.242* *2.056*
Process 3: A & B concentration < 1.5 µg/m³	*6.753*
Process 4: Average A & B concentration > 1000 µg/m³	*0.005*
Process 5: Removing data from sensors with RH issues	*5.527*
Process 6: Removing RH≠ 0-100% and T ≠ 0-130 °F	*3.484*

****percent removed from the total number of observations"***

7- Figure 2 seems to show a wider range of RH with more noise over time. Is this due to seasonal differences or because the RH sensor performance is changing over time?

Response: We thank the reviewer for the comment. However, we did not find an appreciable difference in the RH measurements among the 3 years.

- **2021: Mean RH of 55.07%, range of 20.20 to 80.56%**
- **2022: Mean RH of 54.31%, range of 20.37 to 89.59%**
- **2023: Mean RH of 54.91%, range of 16.43 to 95.04%**

The wider range impression may be illustrated by the fact that January 2021 exhibited a narrower range. A shorter range was also observed for January 2022 and January 2023.

8- Did you consider whether sensor age had any impacts on your results? (e.g., deSouza https://pubs.rsc.org/en/content/articlehtml/2023/ea/d2ea00142j)

Response: We thank the reviewer for this comment. The PurpleAir database did not contain information about the sensors' age or service length. We emphasize the limitations related to the sensors' age in the Results and Discussion section and how that could affect the performance of a model (lines 286-287).

***"The model would, however, be further improved with use of newer PurpleAir sensors because, over time, the quality of the sensors degrades. This is particularly true in the hot and humid climate zone (deSouza et al., 2023)."* (lines 286-287)**

9- "However, DP was excluded as a predictor in our study, because collinearity was found between DP, RH, and T when testing for variance inflation factor. This collinearity is attributed to the direct physical relationship between RH and T" I don't understand what this is saying? T and RH weren't significantly colinear?

Response: We thank the reviewer for this comment. We rephrased the paragraph to make the statement more clear (lines 166-169). RH and T were not collinear. A negative correlation of 14% was found between them. We intended to say that DP was correlated with both RH and T.

"However, DP was excluded as a predictor in our study. DP exhibited collinearity with both RH and T when testing for variance inflation factor. In fact, a high correlation of 95% was found between DP and T. Therefore, including it would inflate the goodness of fit of the model. This result is not surprising considering the interdependent atmospheric thermodynamic relationship of DP with RH and T." **(lines 166-169)**

10- Random withholding is likely not a good test of your model. It would likely be fairer to withhold by site or state. I think it isn't surprising that the model you built for your dataset is a better fit than a model built on another dataset. This is likely something to mention in the limitations.

Response: We appreciate the suggestion. In addition to leave one group out cross-validation (LOGOCV), which leaves out a randomly selected group, we added a leave-one-state-out cross-validation (LOSOCV) process (lines 224-226; 299-300; 357-359) which leaves out one U.S. state in the Southeast U.S. domain at a time.

"Then, we applied a leave-one-state-out cross-validation (LOSOCV) that involves splitting the dataset into specific states to evaluate the performance of the model. In our LOSOCV, every U.S. state was left out successively and used in a validation test, while the remaining states were used to train the model." **(lines 224-226)**

"The LOSOCV resulted in a RMSE and a MAE of 3.31 µg m^{-3} and 2.29 µg m^{-3} respectively for Model 4. These values were higher than those for the LOGOCV process, which is not surprising considering the variability between states." **(lines 299-300)**

"LOSOCV for SSC showed improved performance on average compared with the same process for Model 4 (Table S8), with every state exhibiting lower error metrics than the EPA's target value (≤ 7 µg m^{-3}) for RMSE." **(lines 357-359)**

11- Table 2 this is interesting basically if the RH is high add 5 ug/m3 to the concentration? This difference doesn't seem to be reflected in Figure 6. Is there a typo?

Response: We thank the reviewer for the interesting observation. However, the difference between the 2 models did not only affect the intercept since the RH coefficient is about 10 times greater in Cluster 2 than Cluster 1. We added a sentence in the Results section to highlight the difference (lines 348-350). Moreover, the difference between the two models is not reflected in Figure 6 because Figure 6 shows the correlation between the predicted concentrations after applying the model and AQS concentrations. It would have been noticeable in a figure displaying the relationship between the raw PurpleAir data and the predicted concentrations.

"The difference between the two models resides primarily in their intercepts and their RH coefficients (Table 2). The RH factor is 10 times greater in Cluster 2 than Cluster 1, and the intercept of Cluster 2 is about 5.5 µg m^{-3} greater than Cluster 1." **(lines 348-350)**

12- Citations should be checked for accuracy throughout see a few specific comments below.

Response: We thank the reviewer for pointing out some errors in the citations. They have been corrected.

13- While the results are significantly different, they are not largely different. You might consider adding evaluation of performance by AQI category to further strengthen your findings (e.g., https://www.mdpi.com/1424-8220/22/24/9669, https://amt.copernicus.org/articles/15/3315/2022/amt-15-3315-2022.html)

Response: We appreciate the reviewer's suggestion. We presented and discussed the evaluation of performance of the models by AQI category in the Supplemental Information (Tables S11 and S12). This has been referenced in the manuscript in lines 410-411. Text describing the contents of Table S12 is also included in lines 146-150 of the Supplemental Information:

"Table S12 shows the total percentage of correct AQI reported by each model with their under and over estimation. Models 4 and SSC reported the highest percentage of correct AQIs with a fairly even distribution of under- and overestimation shown by SSC. Model Bj displayed a much higher underestimation than overestimation."

"Table S12: Summary table of the evaluation of the AQI per model for the daily dataset"

Models	*Correct AQI (%)*	*Under-estimation (%)*	*Over-estimation (%)*
SSC	*84.01*	*7.49*	*8.17*
Model 4	*84.10*	*8.70*	*6.87*
Model Bj	*83.78*	*12.81*	*3.07*
Raw PA	*72.68*	*2.99*	*26.94*

Minor

14- A study conducted in 2016 (AQ-SPEC, 2016), evaluating about twelve low-cost PM2.5 sensors showed an overall good agreement between PM2.5 PurpleAir sensors and two reference monitors with a R2 of 78 % and 90 % (Wallace et al., 2021). - Is this citation correct? It seems like the beginning and ending of the sentence are citing 2 different things.

Response: We appreciate the comment. We found this mistake in the bibliography library. The error has been corrected. See lines 42-45.

"A study conducted in 2016 (AQ-SPEC, 2016) to evaluate low-cost $PM_{2.5}$ sensors showed an overall good agreement between $PM_{2.5}$ PurpleAir sensors and two reference monitors with R^2 of 78% and 90% (AQ-SPEC, 2016). However, an overestimation of 40% was found for PurpleAir $PM_{2.5}$ concentrations compared with the reference monitors (AQ-SPEC, 2016; Wallace et al., 2021)." **(lines 42-45)**

15- Lunden, M. M.; Parworth, C. L.; Barkjohn, K. K.; Holder, A. L.; Frederick, S. G.; Clements, A. L. Correction and Accuracy of PurpleAir PM 2.5 Measurements for Extreme Wildfire Smoke. 2022. https://doi.org/10.3390/s22249669. – This citation is incorrect

Response: Thank you for the comment. This was again a mistake in the bibliographic library. It has been corrected to Barkjohn et al. 2022. See lines 242 and the corresponding reference in the bibliography.

16- Line 45, 269: Why are there superscript numbers? Check for this throughout

Response: We appreciate the comment. Line 45 was an error from a change of bibliography style and line 269 was a footnote, for which the corresponded description was missing. All the errors have been corrected. Note that these referenced line numbers are from the last draft.

17- Figure 1: Is the number of counties by state relevant to the story you are telling?

Response: Thank you for the comment. The table with the counties has been removed from Fig. 1.

18- Line 270: "For all the four fitted models, average concentration of 8.80 µg m-3 , with an SD varying between 4.71- 4.84 µg m-3 were obtained, whereas Model Bj provided and a higher MAE than the four developed models with a mean of 7.67 µg m-3 and a SD of 6.08 µg m-3 ." -A little unclear if the first and second part of this sentence are comparing the same thing.

Response: We appreciate the comment. The sentence was restated and reorganized to make the comparison more comprehensible. See lines 277-279 and lines 283-284.

"All four MLR-fitted models exhibited an average concentration of 8.80 µg m^{-3}, with a SD varying between 4.71- 4.84 µg m^{-3}. The Barkjohn model had a mean of 7.67 µg m^{-3} and a SD of 6.08 µg m^{-3}." **(lines 277-279)**

***"The Barkjohn model resulted in a higher MAE than the four models developed for this study."* (lines 283-284)**

19- "Zheng et al. (Zheng et al., 2018) found an R2 280 value of 66 % for a 1-h averaging period after applying an MLR calibration equation to compare three PA sensors" – This is not a paper about PurpleAirs it is a paper about custom built Plantower PMS3003 sensors

Response: We appreciate the comment. We edited the sentence to reflect the specific type of sensor used by Zheng et al. (2018), which is the same type of sensors found within PurpleAir (lines 303-304).

***"Zheng et al. (2018), evaluating the performance of Plantower PMS3003, which is similar to the $PM_{2.5}$ sensor used in PurpleAir,…"* (lines 303-304)**

20- I don't think R2 is usually reported as a Percentage?

Response: Thank you for the comment. R^2 quantifies how much the dependent variable is determined by the independent variables, in terms of proportion of variance. Its values can be presented either in a range from 0 to 1 or in percent. See Wallace et al. (2021) as an example of R^2 stated in %.

21- What is R in Table 1? Just the root of R2?

Response: Thank you for the comment. R is the Pearson correlation. It was defined in line 264.

22- Figure 4: I think this plot would be easier to interpret if both plots used the same color scale.

Response: Thank you for the comment. We used different colors to differentiate between our model and the Barkjohn model. We also added the color description in the figure's caption to avoid confusion.

23- This is a personal preference so take or leave, but I would always put the monitor on the X axis and the Sensor on the Y since the monitor is the independent variable. This is also the recommendation in the EPA performance targets.

Response: We thank the reviewer for expressing this concern. AQS concentrations are shown on the y-axis because they are treated as the dependent variable in the model so that the PurpleAir sensor data can be adjusted accordingly.

24- Figure 7: Is there an assumed T and RH for the lines on this plot?

Response: We thank the reviewer for this comment. However, there is not an assumed T or RH for the lines in the plot. T and RH are used to fit the PA data in the MLR and SSC, however. Fig. 7 is a correlation plot. We clarified in the caption:

"Figure 7: Correlations and regression lines between daily AQS and daily raw/predicted $PM_{2.5}$ concentrations using the MLR, the SSC and Model Bj."